



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

중간 사건이 발생한 자료에서  
생존함수의 비교



연세대학교 대학원  
의학전산통계학협동과정  
의학통계학전공  
하 성 준

# 중간 사건이 발생한 자료에서 생존함수의 비교

지도교수 남 정 모

이 논문을 석사 학위논문으로 제출함



2015년 12월 일

연세대학교 대학원

의학전산통계학협동과정

의학통계학전공

하 성 준

# 하성준의 석사 학위논문을 인준함

심사위원 남 정 모 인

심사위원 정 인 경 인

심사위원 송 기 준 인



연세대학교 대학원

2015년 12월 일

## 감사의 글

설렘과 두려움을 안고 대학원에 입학했던 날이 바로 엊그제 같은데 벌써 2년이라는 시간이 흘러 학생이라는 타이틀을 놓아야 할 때가 되었다는 것이 믿기지 않습니다. 부족한 제가 이곳에서 많은 것을 경험하고 배우면서 성장할 수 있게 도와주신 많은 분들에게 감사의 마음을 전하고자 합니다.

먼저 바쁘신 와중에도 논문을 지도해 주시고 아낌없는 조언과 애정을 주신 남정모 교수님께 깊은 감사를 올립니다. 또한 인자하신 미소로 용기와 격려를 보내주신 정인경 교수님, 세심하게 살펴주시고 다양한 지식을 전해주신 송기준 교수님께 진심으로 감사드립니다. 학부시절은 물론 의학통계학이라는 분야에 발을 내디딜 수 있게 도움을 주신 황창하 교수님께도 깊이 감사드립니다.

그리고 익숙하지 않은 대학원 생활에 잘 적응할 수 있게 도와주신 선배님들, 항상 옆에서 힘이 되어준 순규, 책임감 있는 모습으로 귀감이 된 세휘, 활발하고 유쾌한 세영이, 새로운 길을 찾고 있는 새미, 항상 살갑게 반겨주던 지유, 많이 챙겨주지 못해 미안한 마음이 더 큰 세정이, 소소한 웃음을 안겨주었던 소정이, 묵묵히 본인들의 몫을 잘 해내고 있는 경민이, 지윤이, 명균이, 많은 시간을 함께 보내지 못해 아쉬운 지현이와 해린이에게도 감사의 마음을 전합니다.

또한 언제나 진심어린 마음으로 조언을 아끼지 않은 호영누나, 승배형, 힘들고 지칠 때 제 옆을 지켜주던 석모와 채형이, 항상 든든한 버팀목이 되어준 우리 FC SAS 회원들에게도 감사드립니다.

마지막으로 항상 아낌없는 사랑을 주시고 제 결정을 믿고 응원해 주시는 부모님에게도 사랑하고 감사한 마음을 전합니다.

2015년 12월

하성준 올림

# 차례

표 차례	ii
그림 차례	ii
국문 요약	iii
제1장 서론	1
1.1. 연구 배경 및 목적	1
1.2. 연구 내용 및 방법	2
1.3. 논문의 구성	3
제2장 이론적 배경	4
2.1. 표기법 및 결합밀도함수	4
2.2. 단일 모집단 모형	5
2.3. Landmark method	6
2.4. 로그 순위 검정	7
2.5. Cox의 비례위험모형	8
2.6. Multipoint landmark method	9
제3장 모의실험	11
3.1. 모의실험 설계	11
3.2. 모의실험 결과	12
제4장 실제 자료 적용	21
제5장 결론	24
참고 문헌	26
영문 요약	28

## 표 차 례

표 1. landmark time이 1사분위수와 2사분위수인 경우 각 방법 에 대한 제 1종 오류와 검정력	15
표 2. landmark time이 33백분위수와 66백분위수인 경우 각 방 법에 대한 제 1종 오류와 검정력	16
표 3. Landmark time이 1사분위수, 2사분위수, 3사분위수인 경 우 각 방법에 대한 제 1종 오류와 검정력	17
표 4. Landmark time이 1사분위수와 2사분위수일 때 각 landmark time에서 반응군과 비반응군의 표본수	18
표 5. Landmark time이 33백분위수와 66백분위수일 때 각 landmark time에서 반응군과 비반응군의 표본수	19
표 6. Landmark time이 1사분위수, 2사분위수, 3사분위수일 때 각 landmark time에서 반응군과 비반응군의 표본수	20
표 7. 각 landmark time에서 실제 자료에 대한 5년 생존율과 p 값	23

## 그 림 차 례

그림 1. 대장암 환자 중 이차암이 발생한 군과 발생하지 않은 군의 생존함수	22
그림 2. Landmark method를 적용했을 때 대장암 환자 중 이차 암이 발생한 군과 발생하지 않은 군의 생존함수	23

## 국 문 요 약

### 중간 사건이 발생한 자료에서 생존함수의 비교

임상 분야에서 치료의 효과는 특정 사건에 반응하는 집단과 그렇지 않은 집단의 생존함수의 비교를 통해서 알아보는 경우가 많다. 임상시험에 등록된 환자들은 관찰기간 동안에 투약, 장기 이식 등의 중간 사건을 경험하며 이러한 중간 사건은 생존 함수의 변화를 유발할 수 있다. 예를 들어 항암제 임상시험에서 반응군과 비반응군의 전체 생존기간을 비교하는 경우 항암제로 인해 종양이 사라졌거나 줄어들었다면 환자의 생존 함수가 유리한 쪽으로 바뀌었다고 생각할 수 있다. 그러나 이는 종양 반응을 평가하는 시점보다 환자들이 더 오래 생존해야만 종양 반응을 평가받을 수 있는 왼쪽 절단된 자료이다. 이러한 길이 편의를 고려하여야 옳은 통계적 결과를 얻을 수 있다. 이에 Anderson, Cain, and Gelber(1983)은 임의의 고정된 시점 landmark time을 설정하여 길이 편의를 제거하는 landmark method를 제안하였다.

본 연구에서는 다양한 상황을 가정한 모의실험을 통해 중간 사건이 생존 함수에 어떤 영향을 미치는지와 landmark time이 바뀔 때 따라 결과가 어떻게 바뀌는지 알아보고, 실제 자료에 적용해 보았다. 또한 여러 landmark time에서의 통계량을 하나로 합친 multipoint landmark method와 함께 Cox의 비례위험모형을 사용했을 때의 결과도 함께 알아보았다. 여러 상황에서 시간 종속 공변량을 고려한 Cox의 비례위험모형의 성능이 가장 좋은 것으로 나타났으며, landmark method에서는 landmark time이 하나일 때보다 두 개 이상일 때의 성능이 더 좋은 것으로 나타났다.

---

핵심되는 말 : 생존 분석, landmark method, 중간 사건



# 제1장 서론

## 1.1 연구 배경 및 목적

임상 분야에서 치료의 효과는 특정 사건에 반응하는 집단과 그렇지 않은 집단의 생존 함수의 비교를 통해서 알아보는 경우가 많다. 임상시험에 등록된 환자는 추적 관찰 기간 동안 중간 사건을 경험할 수 있다. 투약, 수술, 장기 이식 등도 중간 사건의 예라고 할 수 있으며, 이러한 중간 사건은 생존 함수의 변화를 유발할 수 있다. 항암제 임상 시험에 등록된 환자들의 종양이 추적 관찰 기간 동안 크기가 줄어들거나 사라졌다면 이는 그 환자의 생존 함수가 유리한 방향으로 바뀌었다고 볼 수 있을 것이다. 반대로 종양의 크기가 커졌거나 환자가 사망했다면 생존 함수는 불리한 방향으로 바뀌었다고 할 수 있을 것이다.

항암제 임상 시험에서 항암제의 효과는 주로 종양 반응이나 전체 생존기간을 통해 평가된다. 환자들이 종양 반응에 의해 반응군과 비반응군으로 나누어졌다면 이 두 군의 생존 함수를 비교하여 항암제의 효과를 평가하는 것이 일반적이다. 그러나 자료에 존재하는 길이 편의를 고려하여야 항암제의 효과를 올바르게 평가할 수 있다. 환자들은 종양 반응을 평가하는 시점보다 더 오래 생존해야만 종양 반응을 평가받을 수 있기 때문이다. 종양 반응을 평가하는 시점 이전에 환자가 사망한다면 이 환자는 비반응군에 속하게 될 것이고, 이는 왼쪽 절단된 자료라고 할 수 있을 것이다. 자료의 길이 편의를 고려하지 않고 두 군의 생존 함수를 비교한다면 잘못된 결과를 도출할 수 있다.

이에 임의의 고정된 시점 landmark time을 설정함으로써 길이 편의를 제거하여 반응군과 비반응군의 올바른 비교를 가능하게 하는 landmark method가 제시되었다 (Anderson, Cain, and Gelber 1983).

본 연구에서 사용한 실제 자료는 1995년 1월 1일부터 2006년 12월 31일까지 연세대학교 세브란스병원 암센터 종양등록 사업소에 등록된 대장암 환자와 대장암 진단 후

이차성 원발암을 진단받은 환자들에 대해 후향적으로 연구한 Yang (2013)의 연구에 사용된 자료이다. 이 자료에는 이차암에 걸린 환자군과 이차암에 걸리지 않은 환자군의 생존 기간이 포함되어 있다. 이차암이라는 중간 사건을 고려하지 않고 분석했을 때에는 이차암에 걸린 군의 생존 기간이 이차암에 걸리지 않은 군에 비해 6개월 더 길었다. 하지만 중간 사건을 고려하여 landmark method를 적용시켰을 때에는 이차암에 걸리지 않은 군의 생존 기간이 더 길게 나타났다. 이는 중간 사건 때문에 생긴 길이 편의를 분석에 고려하였는지의 차이라고 생각할 수 있다.

본 연구에서는 중간 사건이 생존 함수에 어떤 영향을 미치는지와 landmark method에서 landmark time이 바뀔 때 따라 각 상황에서의 결과가 어떻게 바뀌는지 알아보고자 한다. 또한 여러 landmark time에서의 통계량을 하나로 합친 multipoint landmark method와 함께 Cox의 비례위험모형을 사용했을 때의 결과도 함께 알아본다.

## 1.2 연구 내용 및 방법

본 연구에서는 Anderson, Cain, and Gelber (1983)와 Nam and Zelen (2001) 연구의 기본 개념을 이용하기로 한다. 먼저 반응군과 비반응군의 길이 편의를 고려하지 않았을 때의 문제점을 소개하고, landmark method에 대해 소개하고자 한다.

일반적으로 두 군의 생존함수를 비교할 때에는 로그 순위 검정과 Cox의 비례위험모형을 주로 사용한다. 길이 편의를 고려하지 않고 로그 순위 검정을 했을 때의 결과와 시간 종속 공변량을 이용한 Cox의 비례위험모형, 그리고 landmark method를 통한 로그 순위 검정 결과의 차이를 알아본다. 또한 landmark time이 하나가 아닐 때 각 landmark time에서 구해진 Cox의 비례위험모형의 회귀계수들을 하나로 합친 multipoint landmark method도 고려하기로 한다.

중간 사건 전후의 생존 함수, landmark time, 표본수의 여러 상황을 가정한 생존 자료를 생성하여 모의실험을 통해 각 상황에서의 결과가 어떤 차이점이 있는지 살펴보고자 한다. 또한, 실제 자료에 적용시켰을 때에는 어떤 결과가 도출되는지도 알아

보고자 한다.

### 1.3 논문의 구성

제 1장에서는 연구의 배경과 목적, 내용 및 방법을 소개한다. 2장에서는 landmark method, 로그 순위 검정, Cox의 비례위험모형, multipoint landmark method에 관한 이론적 배경과 개념을 정리한다. 3장에서는 모의실험을 통해 각 상황에서 2장에서 소개한 방법들의 성능을 비교해본다. 4장에서는 네 가지 검정 방법을 실제 자료에 적용시켰을 때의 결과를 비교한다. 마지막으로 5장에서는 결론 및 고찰을 제시한다.



## 제2장 이론적 배경

### 2.1 표기법 및 결합밀도함수

본 연구에서는 중간 사건의 형태가 이분형 자료인 경우만을 고려한다. 확률변수  $W$ 와  $T$ 가 각각 중간 사건과 사망이 일어나기까지 걸린 시간이라고 하자. 중간 사건을 나타내는 이분형 확률변수  $Z$ 는  $Z = I(W \leq T)$ 로 정의한다. 중간 사건이 발생하였다면 이분형 확률변수  $Z$ 의 값은 1, 중간 사건이 발생하지 않았다면  $Z$ 의 값은 0이 된다. 또한 중간 사건이 발생하였을 때 사망시간은  $T_1$ , 중간 사건이 발생하지 않았을 때의 사망시간은  $T_0$ 로 정의한다. 사망시간  $T$ 는  $T = (1 - Z)T_0 + ZT_1$ 로 정의할 수 있다.

$W$ ,  $T_0$ ,  $T_1$ 의 확률밀도함수는  $g(w)$ ,  $q_0(t)$ 와  $q_1(t)$ 라 표기하고, 각각의 생존 함수는  $G(w) = \Pr[W > w]$ ,  $Q_0(t) = \Pr[T_0 > t]$ 와  $Q_1(t) = \Pr[T_1 > t]$ 로 정의한다.

만약  $Z = 1$ 이라면 중간 사건이 관측되었다는 의미이고 이를 밀도함수의 형태로 표현한다면 다음과 같다.

$$f(w, z = 1) = Q_0(w)g(w)$$

이를 이용하여 중간 사건이 일어날 확률  $\theta$ 를 구할 수 있다.

$$\theta = P[Z = 1] = \int_0^{\infty} Q_0(w)g(w)dw$$

중간 사건을 경험하지 않고 사망에 이르게 되면 이 관측치는 우중도절단된 자료이다. 이 경우의 결합 확률밀도함수는 다음과 같다.

$$f(t, z=0) = q_0(t)G(t)$$

만약  $T > W$ 라면 사망이 중간 사건 이후에 발생한 것이므로  $z=1$ ,  $T=T_1$ 가 되고 중간 사건이 일어나는 시점  $w$ 는  $w < T_0$  시점에서 관측되고 이 때 생존 함수의 변화가 일어났을 것이라고 생각할 수 있다. 이 경우의 밀도함수는 다음과 같다.

$$f(t, w, z=1) = Q_0(w)g(w)\frac{q_1(t)}{Q_1(w)} \quad \text{for } 0 < w \leq t < \infty$$

만약 중간 사건을 경험한 환자가 사망했다면 이는 중간 사건 이후에 사망이 발생하여야함을 의미한다. 이를 통해 중간 사건이 일어나는 시점까지 살아있어야만 중간 사건을 경험할 수 있는 길이 편의가 존재한다는 것을 알 수 있다. 따라서 생존 시간  $T$ 는 중간 사건의 발생 여부에 따라 결정된다고 할 수 있다. 두 경우에 대하여 조건부 밀도함수를 구하면

$$f(t|z=0) = \frac{q_0(t)G(t)}{\int_0^\infty q_0(t)G(t)dt}$$

$$f(t|z=1) = \frac{A(t)q_1(t)}{\int_0^\infty A(t)q_1(t)dt}$$

이고, 여기서  $A(t) = \int_0^t [Q_0(w)g(w)/Q_1(w)]dw$  이다.

## 2.2 단일 모집단 모형

만약 중간 사건이 생존함수의 변화를 유발하지 않는다면, 중간 사건 여부에 따른

두 조건부 밀도함수 또한 같을 것이라고 생각해볼 수 있다. 즉,  $q_0(t) = q_1(t)$ 이 성립한다면  $f(t|z=0) = f(t|z=1)$  또한 성립할 것이고, 이 때  $A(t)$ 는 다음과 같이 다시 표현할 수 있다.

$$\begin{aligned} A(t) &= \int_0^t \frac{Q_0(w)g(w)}{Q_0(w)} dw \\ &= \int_0^t g(w) dw \\ &= 1 - G(t) \end{aligned}$$

$q_0(t) = q_1(t)$ 라는 가정 하에  $f(z|t=0)$ 와  $f(z|t=1)$ 를 다시 표현하면 다음과 같다.

$$\begin{aligned} f(t|z=0) &= \frac{q_0(t)G(t)}{\int_0^\infty q_0(t)G(t)dt} \\ f(t|z=1) &= \frac{q_0(t)(1-G(t))}{\int_0^\infty q_0(t)(1-G(t))dt} \end{aligned}$$

여기서  $q_0(t) = q_1(t)$ 일 때에  $G(t) \neq (1-G(t))$ 이므로  $f(t|z=0)$ 와  $f(t|z=1)$ 가 같지 않음을 알 수 있다. 따라서 단순히 중간 사건의 경험 여부에 따라서 두 군의 생존함수를 비교하는 것은 잘못된 방법이라고 할 수 있다.

## 2.3 Landmark method

Landmark method에서는 임의의 landmark time  $t_0$ 를 설정하고, 이분형 확률변수  $z(t_0)$ 는  $z(t_0) = I(W < t_0 | T > t_0)$ 로 정의한다. 즉,  $t_0$  이후에 사망한 사람들 중  $t_0$  이전에 중간 사건을 경험했다면  $z(t_0) = 1$ 이 된다. 마찬가지로  $z(t_0) = 0$ 은  $t_0$  이후에 사

망했지만  $t_0$  이전에 중간 사건을 경험하지 못한 경우가 해당된다.  $z(t_0)$ 에 대해서 조건부 밀도함수를 구하면,

$$f(t|z(t_0)=0) = \frac{q_1(t)[A(t)-A(t_0)] + G(t)q_0(t)}{\int_{t_0}^{\infty} \{q_1(x)[A(x)-A(t_0)] + G(x)q_0(x)\}dx} \quad \text{for } t > t_0$$

$$f(t|z(t_0)=1) = \frac{q_1(t)}{Q_1(t_0)} \quad \text{for } t > t_0$$

이고, 만약  $q_0(t) = q_1(t)$ 라면  $f(t|z(t_0)=0) = f(t|z(t_0)=1) = q_0(t)/Q_0(t_0)$ 로 두 조건부 밀도함수가 같아짐을 알 수 있다.

Landmark method를 통해 두 군간 생존함수의 올바른 비교는 가능해졌다. 하지만 landmark method의 분석 결과는 연구자가 임의로 설정한 landmark time  $t_0$ 에 의해 결정된다는 단점이 존재한다. 자료가 모두 수집된 후에 landmark time을 설정하는 것은 또 다른 편의를 유발할 수 있으므로 임상적 의미가 있는 시점을 사전에 landmark time으로 설정해 놓는 것이 좋은 방법이라고 생각된다. Landmark method의 또 다른 단점은 landmark time 이전에 사망한 환자들은 자료 분석에서 제외된다는 것이다. 따라서 landmark time을 길게 설정할 경우 자료의 손실 또한 많이 생긴다는 것도 고려하여야 한다.

## 2.4 로그 순위 검정(Log rank test)

군 간 생존함수의 동질성 여부를 검정할 수 있는 방법은 여러 가지가 있다. 대표적으로 사용하는 방법이 로그 순위 검정(log rank test)이다.  $t_1 < t_2 < \dots < t_D$ 가 서로 다른 사건 발생 시간이라고 하자.  $j = 1, 2, \dots, K$ ,  $i = 1, 2, \dots, D$ 일 때,  $d_{ij}$ 는  $t_i$  시점에서 발생한  $j$ 번째 군의 사건의 수,  $Y_{ij}$ 는  $t_i$  시점에서  $j$ 번째 군의 위험에 노출된 개체

의 수를 나타낸다. 또한  $d_i = \sum_{j=1}^K d_{ij}$ 와  $Y_i = \sum_{j=1}^K Y_{ij}$ 는 각각  $t_i$  시점에서의 사건 수와 위험에 노출된 개체의 수를 나타낸다.  $K=2$ 일 때 로그 순위 검정통계량은 다음과 같다.

$$Z = \frac{\sum_{i=1}^D W(t_i) \left[ d_{i1} - Y_{i1} \left( \frac{d_i}{Y_i} \right) \right]}{\sqrt{\sum_{i=1}^D W(t_i)^2 \frac{Y_{i1}}{Y_i} \left( 1 - \frac{Y_{i1}}{Y_i} \right) \left( \frac{Y_i - d_i}{Y_i - 1} \right) d_i}} \sim \chi_1^2$$

모든 시점  $t_i$ 에 대해서 가중치  $W(t_i)$ 가 1인 경우가 일반적인 로그 순위 검정통계량이다. 가중치  $W(t_i)$ 에 따라 다른 형태의 가중 로그 순위 검정통계량이 존재한다.

## 2.5 Cox의 비례위험모형(Cox' s proportional hazard model)

Cox의 비례위험모형(Cox' s proportional hazard model)은 위험함수를 기초로 하는 회귀모형이다. 위험함수(hazard function)는 임의의 시점  $t$ 에서 사건이 발생할 확률을 의미한다. 임의의 시점  $t$ 에서의 확률밀도함수를  $f(t)$ , 생존함수를  $S(t)$ , 누적위험함수를  $H(t)$ 라고 하면, 위험함수를 다음과 같이 생존함수와 확률밀도함수의 관계로 표현할 수 있다.

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

또한 누적위험함수를  $H(t) = \int_0^t h(u)du$ 라 하면 생존함수는 누적위험함수와 관계에 의하여 다음과 같이 표현할 수 있다.



$$S(t) = \exp\left[-\int_0^t h(u)du\right] = \exp[-H(t)]$$

Cox의 비례위험모형은 반응변수인 생존시간과 공변량의 관계가 위험함수를 통해서 표현된다. Cox의 비례위험모형은 모든 독립변수가 비례위험 가정을 만족해야 한다. Cox의 비례위험모형은 다음과 같다.

$$h(t|\mathbf{Z}) = h_0(t)\exp(\boldsymbol{\beta}\mathbf{Z})$$

위 식에서  $h_0(t)$ 는 분포 가정이 주어지지 않은 기저위험함수(baseline hazard function)이며, 이는 공변량 값이 0일 때,  $t$ 시점에서의 위험함수를 말한다.  $\boldsymbol{\beta}'$ 는 공변량  $\mathbf{Z}$ 의 효과를 나타내는 회귀계수를 나타낸다.

Cox의 비례위험모형은 모형에 시간 종속 공변량(time dependent covariates)을 포함할 수 있다. 시간 종속 공변량  $\mathbf{Z}(t)$ 를 이용하여 Cox의 비례위험모형을 표현하면 다음과 같다.

$$h(t|\mathbf{Z}(t)) = h_0(t)\exp(\boldsymbol{\beta}_{TD}\mathbf{Z}(t))$$

## 2.6 Multipoint landmark method

2.3장에서 소개한 landmark method에서는 임의의 한 시점  $t_0$ 에 대해서  $z(t_0)$ 를 정의하였다. Landmark time  $t_0$ 는 연구자의 선택에 따라 달라질 수 있을 것이고 그 시점에서 반응군과 비반응군의 생존함수의 동질성 여부는 로그 순위 검정으로 검정할 수 있을 것이다. 하지만 landmark time  $t_0$ 를 어떻게 설정할 것인가에 대한 문제점이 존재한다. 본 연구에서는 landmark time이 하나가 아닐 때, 즉  $s_i(i=2, \dots, k)$ 일 때 Cox

의 비례위험모형에서 구해진 회귀계수들을 이용한 검정통계량을 고려해보았다.

임의의 landmark time  $s_i$ 에 대하여 Cox의 비례위험모형을 표현하면 아래와 같다.

$$h(t|\mathbf{Z}^{(s_i)}) = h_0(t|s_i)\exp(\beta_{LM}^{(s_i)}\mathbf{Z}^{(s_i)}) \quad \text{for } t > s_i$$

Landmark time  $s_i$ 가 변함에 따라 위험개체수와 회귀계수  $\hat{\beta}_{LM}^{(s_i)}$ 도 달라질 것이다. 각 시점  $s_i$ 에서의 가중치를  $w_i$ 라 했을 때,  $w_i$ 와  $\hat{\beta}_{LM}^{(s_i)}$ 의 가중 평균을 이용하여  $T^*$ 를 정의하였다.

$$T^* = \frac{\sum_{i=1}^k w_i \hat{\beta}_{LM}^{(s_i)}}{\sum_{i=1}^k w_i}$$

$T^*$ 의 표준오차를 구하여 귀무가설의 기각여부를 판단하는 것은 쉽지 않은 문제이므로 않으므로 붓스트랩(bootstrap)방법을 통해 이를 해결하고자 하였다. 각 반복수행 단계에서 생성된 표본으로부터 복원 추출을 통해  $l$ 개의 붓스트랩 표본을 얻는다.  $l$ 개의 붓스트랩 표본마다  $\hat{T}^*$ 을 구한 뒤,  $(\hat{T}_{0.025}^*, \hat{T}_{0.975}^*)$ 를  $\hat{T}^*$ 의 95% 신뢰구간으로 사용하였다. 이 때,  $\hat{T}_{0.025}^*$ 와  $\hat{T}_{0.975}^*$ 는 각각  $\hat{T}^*$ 의 2.5백분위수와 97.5백분위수를 나타낸다. 구해진  $\hat{T}^*$ 의 95% 신뢰구간에 0이 포함되는지를 통해 귀무가설의 기각여부를 판단하였다.

## 제3장 모의실험

### 3.1 모의실험 설계

이 장에서는 일반적인 로그 순위 검정, landmark method를 통한 로그 순위 검정, 시간 종속 공변량을 고려한 Cox의 비례위험모형, 그리고 multipoint landmark method의 성능을 비교한다. 중간 사건 전후의 생존함수가 같다고 가정했을 때 각 상황에서의 1종 오류를 비교한다. 또한 중간 사건 전후의 생존함수가 다르다고 가정했을 때에는 각 상황에서의 검정력을 비교한다. Landmark method에서 landmark time의 개수는 2개와 3개인 경우만 고려하기로 한다. Multipoint landmark method에서는 시점  $s_i$ 에서의 위험개체수를 가중치  $w_i$ 로 사용한 weighted multipoint landmark method와 모든 시점에서  $w_i = 1$ 을 사용한 unweighted multipoint landmark method 두 경우를 비교하였다.

중간 사건과 사망이 일어나기까지 걸린 시간을 각각  $W$ 와  $T$ 라고 하자. 중간 사건이 관측되었을 때의 사망 시간은  $T_0$ , 중간 사건이 관측되지 않았을 때의 사망시간은  $T_1$ 으로 나타낸다.  $W$ ,  $T_0$ ,  $T_1$ 의 확률밀도함수는  $g(w)$ ,  $q_0(t)$ 와  $q_1(t)$ 라고 표기한다. 또한 이분형 확률변수  $Z$ 는  $Z = I(W \leq T)$ 로 정의하였고  $T = (1 - Z)T_0 + ZT_1$ 로 정의하였다.

따라서  $Z = 0$ 인 경우  $T = T_0$ 가 되고,  $Z = 1$ 인 경우  $T = T_1$ 이 된다. 하지만 후자의 경우  $T_1$ 은  $w$ 시점에서 왼쪽 절단되었기 때문에  $T_1$ 의 확률밀도함수를 사용할 수 없다. 이 경우 생존함수를 이용해  $w$ 시점에서의 생존확률  $v$ 를 구하고, 균등분포  $(0, v)$ 에서 생성된 임의의 실수  $u$ 가 생존확률이 되는 시점을  $T$ 로 사용한다.

본 연구에서  $g(w)$ ,  $q_0(t)$ 는 평균이  $\lambda_{\mu_0}$ 인 지수분포를 따른다고 가정하고,  $q_1(t)$ 는 평균이  $\lambda_{\mu_1}$ 인 지수분포를 따른다고 가정한다. 가능한  $\lambda_{\mu_0}$ 와  $\lambda_{\mu_1}$ 는 무수히 많겠지만 본 연구에서는  $\lambda_{\mu_0}$ 는 12,  $\lambda_{\mu_1}$ 는 12, 18, 24, 30, 36인 경우만을 고려하였다. 또한

landmark method에서 landmark time이 2개인 경우에는  $T$ 의 1사분위수와 2사분위수 또는  $T$ 의 33백분위수와 66백분위수를 사용하는 두 가지 방법을 고려하기로 한다. Landmark time이 3개인 경우에는  $T$ 의 1사분위수, 2사분위수, 그리고 3사분위수를 사용하였다.

생성된  $W$ ,  $T_0$ ,  $T_1$ ,  $Z$ ,  $s_i$ 를 이용하여 로그 순위 검정, landmark method를 통한 로그 순위 검정, 시간 종속 공변량을 이용한 Cox의 비례위험모형과 함께 multipoint landmark method의 성능을 1000번 반복 수행을 통해 비교한다. 표본수는 300, 500인 상황으로 고려하였다. Multipoint landmark method에서  $T^*$ 의 신뢰구간을 구하기 위해 붓스트랩 방법을 500번 수행하였다. 구해진 신뢰구간에 0이 포함되는지를 통해 귀무가설의 기각여부를 판단하기로 한다.

또한 landmark time에 따라 자료의 손실이 어느 정도 발생하는지와 함께 각 landmark time에서 반응군과 비반응군으로 분류된 표본의 수도 함께 알아본다.

### 3.2 모의실험 결과

Landmark time이  $T$ 의 1사분위수와 2사분위수인 경우의 모의실험 결과는 표 1에 나타나 있다.  $q_0(t) = q_1(t)$ 라고 가정했을 때, 즉  $\lambda_{\mu_0} = \lambda_{\mu_1} = 12$ 인 경우 중간 사건을 고려하지 않은 로그 순위검정은 제 1종 오류가 매우 높은 것으로 나타났다. 반면에 landmark method와 시간 종속 공변량을 이용한 Cox의 비례위험모형은 제 1종 오류가 5% 정도로 로그 순위검정에 비하여 매우 낮은 것을 확인할 수 있다. 또한  $q_0(t) \neq q_1(t)$ 인 경우, 즉  $\lambda_{\mu_0} = 12$ 이고  $\lambda_{\mu_1} \neq 12$ 인 경우의 검정력은 시간 종속 공변량을 이용한 Cox의 비례위험모형, multipoint landmark method, landmark method 순으로 나타났다. Landmark method에서는 landmark time이 1사분위수일 때의 검정력이 2사분위수일 때보다 조금 더 높았다. Multipoint landmark method에서는 weighted multipoint landmark method의 검정력이 unweighted multipoint landmark method보다 더 높았지만 차이는 크지 않은 것으로 나타났다.

Landmark time이  $T$ 의 33백분위수와 66백사분위수인 경우의 모의실험 결과는 표 2에 나타나 있다. 중간 사건 전과 후의 생존함수가 같다고 가정했을 경우에는 표 1과 마찬가지로 로그 순위검정의 제 1종 오류가 매우 높게 나타났고, 나머지 방법들은 5% 정도로 나타났다. 중간 사건 전후의 생존함수가 다르다고 가정했을 때 검정력은 시간 종속 공변량을 이용한 Cox의 비례위험모형이 가장 높았고, landmark time을  $T$ 의 33백분위수로 사용했을 때가 그 뒤를 이었다. 이 경우 landmark time을  $T$ 의 66백분위수로 사용했을 때와 multipoint landmark method는 표 1의 결과에 비해 낮은 검정력을 보였다.

Landmark time이 다른 두 상황에서 landmark method를 비교해보면 landmark time이  $s_1$ 일 때 두 결과의 큰 차이는 없었다. 하지만 landmark time이  $s_2$ 일 때, 즉  $T$ 의 2사분위수일 때와 66백분위수일 때의 검정력의 차이는 크게 나타났다. Landmark time이  $T$ 의 66백분위수일 때의 검정력이  $T$ 의 2사분위수일 때에 비해 많이 떨어지는 것을 확인할 수 있었다. 이는 landmark time이 뒤쪽으로 갈수록 분석에서 제외되는 표본이 많아졌고, 중간 사건을 경험한 군과 그렇지 않은 군의 표본수의 불균형 또한 심해졌기 때문이라고 생각된다. 또한 대부분의 상황에서 weighted multipoint landmark method가 unweighted multipoint landmark method보다 더 좋은 성능을 보였다.

Landmark time이  $T$ 의 1사분위수, 2사분위수, 그리고 3사분위수로 3개인 경우의 모의실험 결과는 표 3에 나타나 있다. 이 경우에도 중간 사건 전과 후의 생존함수가 같다고 가정했을 경우 앞서 수행한 모의실험들과 마찬가지로 로그 순위검정의 제 1종 오류가 매우 높게 나타났고, 나머지 방법들은 5% 정도로 나타났다. 중간 사건 전후의 생존함수가 다르다고 가정했을 때 검정력 또한 앞선 결과들과 비슷하게 시간 종속 공변량을 이용한 Cox의 비례위험모형, multipoint landmark method, landmark method 순으로 높았다. Landmark method에서 검정력은 landmark time이  $T$ 의 1사분위수일 때가 가장 높았고, 그 뒤로 2사분위수, 3사분위수 순으로 나타났다. 하지만 Landmark time을  $T$ 의 3사분위수로 설정했을 때에는 검정력이 현저하게 낮은 모습을 보였다. Multipoint landmark method에서는 가중치에 따른 결과의 차이는 크지 않은 것으로 나타났다.

표 4부터 6은 landmark time에 따른 반응군과 비반응군의 표본수를 나타낸다. Landmark time을 뒤쪽으로 설정할수록 반응군의 표본수가 비반응군에 비해 많은 것으로 나타났다. Landmark time이 동일한 상황에서는  $\mu_{\lambda_1}$ 의 값이 커질수록 반응군에 속하는 표본의 수는 많아지고, 비반응군에 속하는 표본의 수는 줄어드는 경향을 보였다. 또한 landmark time이  $T$ 의 50백분위수 이후인 경우 자료의 불균형이 매우 심하게 나타났다.



표 1. landmark time이 1사분위수와 2사분위수인 경우 각 방법에 대한 제 1종 오류와 검정력

Sample size	$\mu_{\lambda_0}$	$\mu_{\lambda_1}$	Log-rank test	Landmark method		Multipoint landmark method		Time dependent Cox model
				$s_1$	$s_2$	Weighted	Unweighted	
300	12	12	100.0	6.6	5.0	4.8	4.6	6.4
		18	100.0	19.3	21.5	24.0	24.5	78.1
		24	100.0	45.8	39.3	54.9	53.7	99.4
		30	100.0	61.7	51.5	71.4	68.7	100.0
		36	100.0	72.3	60.6	80.6	76.8	100.0
500	12	12	100.0	6.8	5.8	5.0	5.3	4.8
		18	100.0	32.8	28.9	39.0	39.5	93.6
		24	100.0	66.7	58.0	76.2	73.4	100.0
		30	100.0	84.3	70.8	90.9	88.8	100.0
		36	100.0	90.1	79.0	95.4	93.7	100.0

표 2. landmark time이 33백분위수와 66백분위수인 경우 각 방법에 대한 제 1종 오류와 검정력

Sample size	$\mu_{\lambda_0}$	$\mu_{\lambda_1}$	Log-rank test	Landmark method		Multipoint landmark method		Time dependent Cox model
				$s_1$	$s_2$	Weighted	Unweighted	
300	12	12	100.0	4.7	4.1	4.1	4.2	8.0
		18	100.0	23.2	12.2	20.5	16.7	82.1
		24	100.0	46.3	27.1	42.9	36.1	100.0
		30	100.0	69.2	33.9	55.1	47.6	100.0
500	12	12	100.0	4.5	6.3	4.6	4.8	4.4
		18	100.0	34.5	18.7	36.3	31.2	93.5
		24	100.0	68.2	33.2	64.5	50.9	99.9
		30	100.0	88.1	39.4	68.7	57.5	100.0



표 3. landmark time이 1사분위수, 2사분위수, 3사분위수인 경우 각 방법에 대한 제 1종 오류와 검정력

Sample size	$\mu_{\lambda_0}$	$\mu_{\lambda_1}$	Log-rank test	Landmark method			Multipoint landmark method		Time dependent Cox model
				$s_1$	$s_2$	$s_3$	Weighted	Unweighted	
1000	12	12	100.0	4.8	4.5	5.0	4.5	5.4	5.2
		18	100.0	56.7	50.3	19.4	69.4	68.5	99.8
		24	100.0	91.4	84.5	28.8	96.6	96.3	100.0
		30	100.0	98.9	93.9	30.4	100.0	100.0	100.0

표 4. landmark time이 1사분위수와 2사분위수일 때 각 landmark time에서 반응군과 비반응군의 표본수

Sample size	$\mu_{\lambda_0}$	$\mu_{\lambda_1}$	$s_1$		$s_2$	
			반응군	비반응군	반응군	비반응군
300	12	12	56.41	168.59	74.79	75.21
		18	61.60	163.40	86.52	63.48
		24	64.11	160.89	93.96	56.04
		30	65.77	159.23	99.41	50.59
		36	67.67	157.33	103.92	46.08
		12	93.71	281.29	125.46	124.54
500	12	18	102.52	272.48	144.38	105.62
		24	106.34	268.66	156.67	93.33
		30	109.58	265.42	166.10	83.89
		36	112.50	262.50	173.31	76.69

표 5. landmark time이 33백분위수와 66백분위수일 때 각 landmark time에서 반응군과 비반응군의 표본수

Sample size	$\mu_{\lambda_0}$	$\mu_{\lambda_1}$	$s_1$		$s_2$	
			반응군	비반응군	반응군	비반응군
300	12	12	66.90	133.10	66.42	33.58
		18	74.51	125.49	77.56	22.44
		24	78.29	121.71	84.42	15.58
		30	81.98	118.02	89.19	10.91
		36	84.29	115.71	92.33	7.96
		12	111.08	221.92	110.98	56.02
500	12	18	123.61	209.39	129.80	37.20
		24	131.25	201.75	141.63	25.37
		30	136.74	196.26	149.34	17.66
		36	140.29	192.71	154.20	12.80

## 제4장 실제 자료 적용

본 연구는 중간 사건이 있는 생존 자료에서 여러 가지 검정 방법의 성능을 비교하는 것이 목적이다. 3장에서는 모의실험을 통하여 검정 방법의 성능을 비교해 보았으며, 4장에서는 실제 임상에서 얻어진 자료를 이용하여 여러 검정 방법의 성능을 비교하고자 한다.

비교에 사용될 실제 자료는 1995년 1월 1일부터 2006년 12월 31일까지 연세대학교 세브란스병원 암센터 종양등록 사업소에 등록된 대장암 환자와 대장암 진단 후 이차성 원발암을 진단받은 환자들에 대해 후향적으로 연구한 Yang (2013)의 연구에 사용된 자료이다. 전체 등록된 환자 5,888명 중 5,487명은 대장암만 있는 환자이며, 나머지 401명은 대장암 진단 후 이차암이 발생한 환자이다. 본 연구에서 사용된 변수는 대장암 진단일, 이차암 발생여부, 이차암 진단일, 사망여부, 사망일이며, 마지막 추적 관찰 날짜인 2012년 5월 1일까지 생존한 환자는 2012년 5월 1일부로 중도절단되었다. 이 자료를 이용하여 이차암 발생여부를 고려했을 때와 고려하지 않았을 때 결과의 차이를 비교하였다. 또한 landmark method와 시간 종속 공변량을 포함한 Cox의 비례위험모형의 결과도 비교하였다.

대장암만 있는 환자군의 중앙 생존 기간은 11년 8개월이고 이차암 발생군의 중앙 생존 기간은 12년 2개월로 나타났다. 또한 두 군의 3년 생존율은 각각 68.0%, 80.8%였으며 5년 생존율은 각각 59.5%, 71.1%로 이차암 발생군의 생존율이 더 높았으며 이는 통계적으로 유의한 차이를 보였다(로그 순위 검정,  $p=0.0078$ ). 그러나 이 결과는 이차암 발생 전에 사망한 환자들 이차암이 없는 군으로 분류되었으므로 길이 편의를 고려하지 못한 결과라고 할 수 있다.

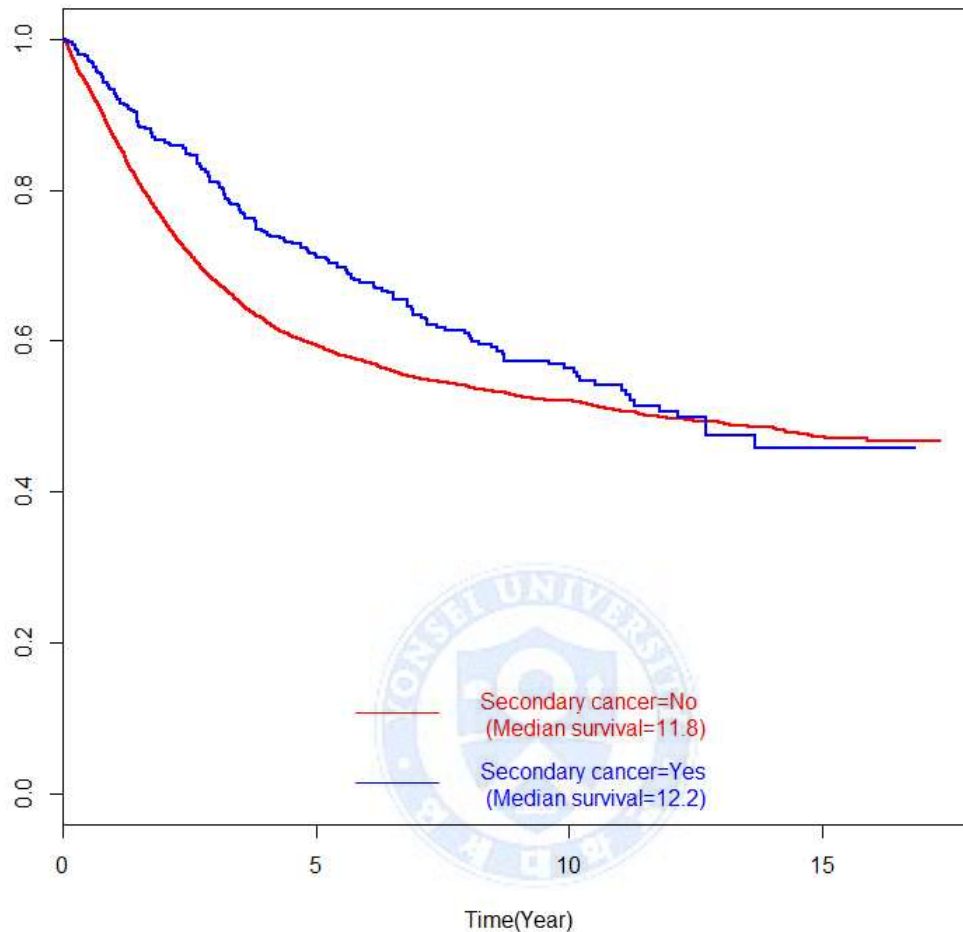


그림 1. 대장암 환자 중 이차암이 발생한 군과 발생하지 않은 군의 생존함수

본 실제 자료를 landmark method에 적용시키기 위하여 대장암 진단일로부터 1년, 3년, 5년 후를 landmark time으로 설정한 뒤 landmark method를 적용하여 각 landmark time에서의 이차암 발생 여부에 따른 생존율을 비교하였다. 세 경우 모두 이차암이 있는 군에서의 생존율이 더 낮은 것으로 나타났다.

같은 자료를 시간 종속 공변량을 고려한 Cox의 비례위험모형에 적용시켰을 때에는 이차암이 발생한 군은 이차암이 발생하지 않은 군에 비해 사망에 대한 위험도(hazard ratio)가 2.765배 더 높은 것을 확인할 수 있었다( $p\text{-value} < 0.0001$ ).

표 7. 각 landmark time에서 실제 자료에 대한 5년 생존율과 p 값

Landmark time	5년 생존율(%)		p 값
	이차암 발생군	이차암 비발생군	
1년	55.0	66.5	0.0009
3년	68.2	79.3	0.0011
5년	74.2	87.3	<0.0001

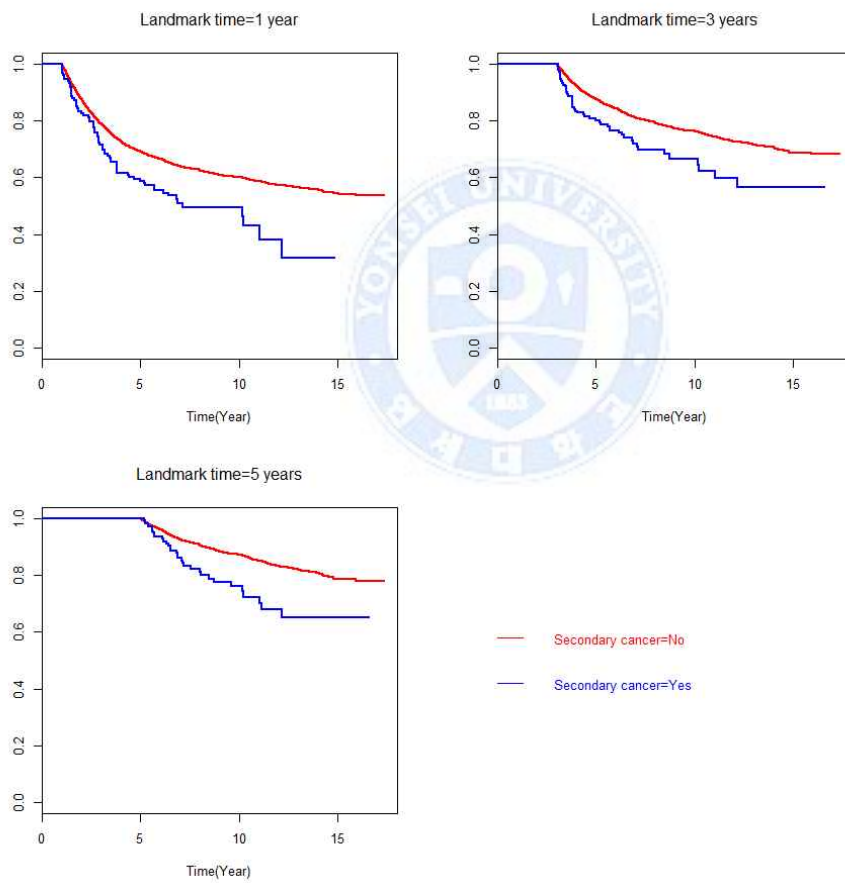


그림 2. Landmark method를 적용했을 때 대장암 환자 중 이차암이 발생한 군과 발생하지 않은 군의 생존함수

## 제5장 결론

임상 분야에서 치료의 효과는 특정 사건에 반응하는 집단과 그렇지 않은 집단의 생존함수의 비교를 통해서 알아보는 경우가 많다. 임상시험에 등록된 환자들은 관찰기간 동안에 투약, 장기 이식 등의 중간 사건을 경험할 수 있다. 중간 사건은 중간 사건이 일어날 때까지의 생존을 필요로 하므로 길이 편이가 발생하게 된다. 이러한 자료의 길이 편이를 고려하지 않고 두 군의 생존 함수를 비교한다면 잘못된 결과를 도출해낼 수 있다.

본 연구에서는 중간 사건이 중간 사건 전후의 생존함수에 변화를 유발한다고 가정했을 때와 유발하지 않는다고 가정했을 때 여러 가지 검정 방법들의 제 1종 오류와 검정력을 모의실험을 통해 비교해 보았고 실제 자료에도 적용시켜 보았다.

모의실험 결과 중간 사건이 생존함수에 변화를 유발하지 않음에도 불구하고 중간 사건을 고려하지 않고 로그 순위 검정을 수행한다면 잘못된 결론을 내릴 수 있다는 것을 확인할 수 있었다. 반면 landmark method와 시간 종속 공변량을 이용한 Cox의 비례위험모형은 제 1종 오류가 5% 내외로 유지되는 성능을 보였다. 또한 중간 사건이 생존함수의 변화를 유발한다고 가정했을 때에는 Cox의 비례위험모형이 가장 좋은 검정력을 보였고, 하나의 landmark time을 사용한 landmark method보다는 두 개의 landmark time을 사용한 multipoint landmark method의 검정력이 대체로 더 좋은 것을 확인할 수 있었다.

실제 자료에서는 이차암의 발생 여부를 고려하지 않을 경우 생기는 길이 편이를 제거하기 위하여 landmark method를 적용해 보았다. Landmark method는 임의의 고정된 시점 landmark time에서 이차암이 발생했는지의 여부에 따라 두 군으로 나누고 landmark time 이전에 사망한 환자는 분석에서 제외시킨 뒤 이차암이 발생한 군과 발생하지 않은 군의 생존 시간을 비교함으로써 이차암 발생 전 사망한 환자가 이차암이 발생하지 않은 군에 포함되는 길이 편이를 제거해주는 방법이다. 대장암 진단 후 1년, 3년, 5년 후를 landmark time으로 설정한 뒤 분석한 결과 이차암이 발생한 군은

이차암이 발생하지 않은 군에 비해 사망에 대한 위험도가 각각 1.501배, 1.612배, 2.014배로 더 높았다.

본 연구에서는 중간 사건으로 인한 길이 편의가 생존 함수의 비교에 어떤 영향을 미치는지와 중간 사건이 존재하는 자료에서 길이 편의를 제거하여 올바른 결과를 도출해내는 방법들을 소개, 비교하였다. 본 연구에서는 임상 시험 자료만을 소개하였으나 임상 시험이 아닌 다른 분야에서도 여러 형태의 중간 사건이 존재할 것이라고 생각된다. 중간 사건으로 인한 길이 편의를 고려하지 않고 분석을 한다면 잘못된 결과를 얻을 수 있으므로 유의하여야 한다. 본 연구에서는 2개와 3개의 landmark time을 이용한 multipoint landmark method만을 고려하였지만 그 이상으로 확장해 보는 것도 가능할 것이라고 생각된다.





## 참고문헌

양지현. 대장암 환자에서 이차성 원발암 발생 관련 요인과 생존율 연구. 연세대학교 대학원 석사학위 논문, 2012년

김양진. 2013. 생존분석. 과주: 자유아카데미.

J. Klein, M. Moeschberger, Survival Analysis Techniques for Censored and Truncated Data, Springer, New York 1997.

Anderson, J. R., Cain, K. C., & Gelber, R. D. (1983). Analysis of survival by tumor response. J Clin Oncol, 1(11), 710-719.

Nam, C. M., & Zelen, M. (2001). Comparing the survival of two groups with an intermediate clinical event. Lifetime Data Anal, 7(1), 5-19.

Lefkopoulou, M., & Zelen, M. (1995). Intermediate clinical events, surrogate markers and survival. Lifetime Data Anal, 1(1), 73-85.

Dafni, U. (2011). Landmark analysis at the 25-year landmark point. Circ Cardiovasc Qual Outcomes, 4(3), 363-371. doi: 10.1161/CIRCOUTCOMES.110.957951

Buyse, M., & Piedbois, P. (1996). On the relationship between response to treatment and survival time. Stat Med, 15(24), 2797-2812. doi: 10.1002/(SICI)1097-0258(19961230)

D. R. Cox (1972). "Regression models and life-tables." Journal of the Royal Statistical Society. Series B (Methodological): 187-220.



## Abstract

### Comparison of survival time with an intermediate event

Ha, Seong Jun

Dept. of Biostatistics and Computing

The Graduate School

Yonsei University

In clinical trials, the effect of therapy is often studied by comparing survival distribution of the groups that respond to specific incidents and the groups that do not respond. Patients registered for clinical trials experience intermediate events such as medication and organ transplant during observational period, and these intermediate events can cause the change of survival distribution. For instance, in comparing responders and nonresponders of overall survival in cancer trials, if a patient's tumor disappears or shrinks due to anticancer drugs, the patient's survival distribution has been changed into favorable condition. However, this is left truncated data that tumor response can be judged when patients should survive longer than the period of evaluating tumor response. The length bias should be considered to acquire correct statistic result. Therefore, Anderson, Cain, and Gelber(1983) suggested landmark method that removes length bias by setting randomly set landmark time.

In this study, through simulation that assumed various situations, what effects intermediate events have on survival distribution was researched along with how the result changed according to changing landmark time. It was applied

to the real data. Besides, we researched the result when Cox' s proportional hazard model was used with multipoint landmark method that many statistics of landmark time were combined. In various situations, the performance of Cox's proportional hazard model with time dependent covariance was the best, and in landmark method, the performance was better with more than two landmark times than one.

---

Key words : survival analysis, landmark method, intermediate event

